HBFC BANK (Personal Loan)

Problem Statement:

This case is about a bank (HBFC Bank) which has a growing customer base. Majority of these customers are customers having deposits (Saving, Current) and Term Deposit (including Fixed Deposits and Recurring deposit accounts) in the bank. The number of customers who are also borrowers (customers having loan accounts) is quite small, and the bank is interested in expanding this base rapidly to bring in more loan business and in the process, earn more through interest on loans. In particular, the management wants to explore ways of converting its customers having deposits to personal loan customers (while retaining them as depositors).

The bank wants to build a model that will help them identify the potential customers who have a higher probability of purchasing the loan. For doing that, the first step in this regard is to explore the previous data and drill insights.

You have been provided with a dataset of 5000 customers. The data include customer demographic information (age, income, etc.), the customer's relationship with the bank (mortgage, securities account, etc.), and the customer response to the last personal loan campaign (Personal Loan).

You are brought in as a consultant and your job is to explore the data to understand the variables and the impact, they have had on Personal Loans so that the bank can leverage the insights to reach out to the right customers who have a higher probability of purchasing the loan.

Question 1:

What percentage of the bank's customers (according to the data) have availed Personal Loans?

Customers who did not availed Personal Loan	4520
Customers who availed Personal Loan	480
Total Customers	5000
Percentage of customers who availed the person	9.6
	10%

Using 'IF' and 'COUNTIF' conditions we calculated the percentage of customers who availed Personal Loans. The answer to this question is 9.6% or almost 10% of the customers availed Personal Loans.

Question 2:

Generate a table with min, max, median & average for all numeric variables (age, experience, income, family members, CCAvg, Mortgage). What are your observations?

Column1	AGE	EXPERIENCE	INCOME	FAMILY MEMBERS	CCAVG	MORTGAGE
MIN	23	0	8	1	0	0
MAX	67	43	224	4	10	635
MEDIAN	45	20	64	2	1.5	0
AVERAGE	45.3384	20.1348	73.7742	2.397230028	1.937938	56.4988

Age: The minimum age in the data is a customer who is 23 years old and the maximum being 67 years old. The Median and Average age is around 45 years, indicating that most of the customers are in their mid 40s.

Experience: The minimum experience is 0 years, and the maximum is 43 years. The Median and Average experience is that of 20 years, indicating that many customers have significant work experience.

Income: With the income range from 8k/year to 224k/year, the range is wide. Half the customers earn less than the median income which is 64k/year and the other earns more than that. At approximately 74k/year, the income range is slightly higher for the customers.

Family Members: Each customer has atleast 1 family member and at most 4 family members, with the median being 2. The Average family size is of 2 indicating that the customers have relatively small families.

CCAvg: The Credit Card spending range is from 0 to 10k/year for the customers. With the median being 1.5k/year half the customers spend more than this per year and the other half spends less than the median. The average amount being almost 2k/year indicates that the customers have relatively low credit card spending.

Mortgage: It ranges from 0 to 635k/year which is very wide. The Median mortgage being 0 indicates that half of the customers do not have any mortgages. The Average is slightly in higher range because few customers have high mortgage value. Therefore, influencing the average.

Question 3:

Create a new categorical variable for Experience using 4 categories –

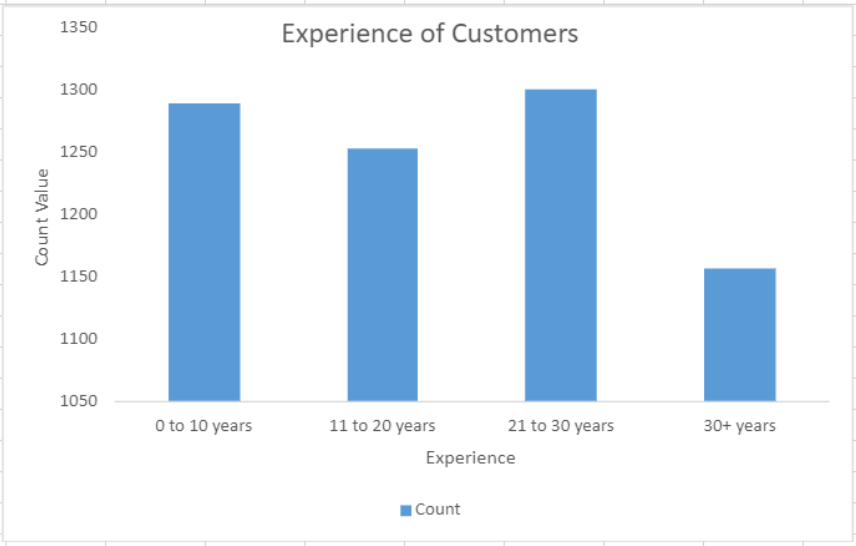
a. 0 to 10 years

b. 11 to 20 years

c. 21 to 30 years

d. 30+ years.

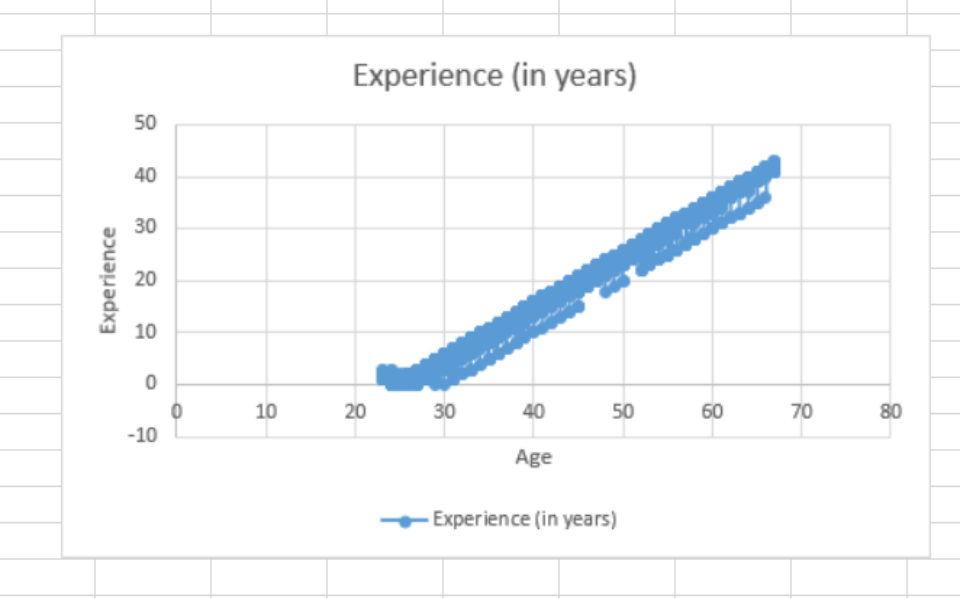
Categories	Count
0 to 10 years	1289
11 to 20 years	1253
21 to 30 years	1301
30+ years	1157
Total	5000



Using 'Nested IF and COUNTIF' conditions a new categorical variable on experience was created. Using this variable a bar plot was plotted.

Question 4:

Create a scatter plot of the Age and the Experience variable. What do you observe?



This scatter plot uses Age and Experience in its axes. The gradual upward growth in the curve indicates positive relationship between Age and Experience. It indicates that as the age increases so does the experience

Question 5:

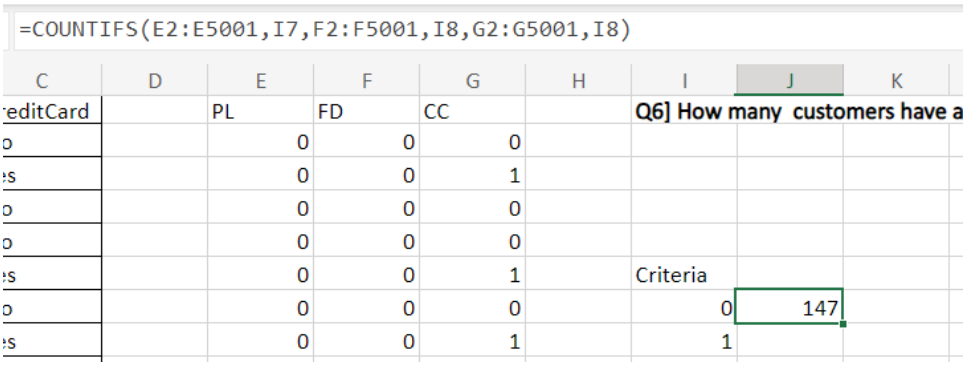
What are the top 3 areas (ZIP Codes) where the bank's customers are located?

Top 3 Zip Codes	Count of top 3 Zip Codes
94720	169
94305	127
95616	116

After finding the unique zip codes and using 'COUNTIF' and 'LARGE' conditions to check the frequency of the code and the count, we found the top 3 areas where the bank customers are located.

Question 6:

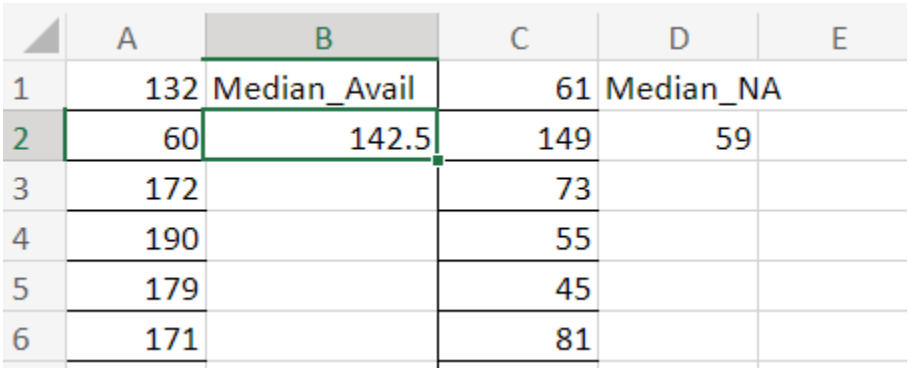
How many customers have a combination of Fixed Deposits and Credit Cards but not Personal Loan?



Using ‘IF’ condition we found which customer has Fixed Deposits, Credit Cards or has availed Personal Loan. Then using ‘COUNTIF’ condition we got the count of the customers who have a combination of Fixed Deposits and Credit cards. The answer is 147 customers have this combination.

Question 7:

What is the median income of the customers who have availed personal loans and compare it with the median income of those customers who have not availed personal loans? What do you infer?



	A	B	C	D	E
1	132	Median_Avail	61	Median_NA	
2	60	142.5	149	59	
3	172		73		
4	190		55		
5	179		45		
6	171		81		

The median salary of customers who availed of Personal Loan is approximately 143k/year.

The median salary of customers who did not avail of Personal is 59k/year.

Inference: Customers with higher income range have high median range as compared to that of customers with lower income range.

Question 8:

Create 4 separate Pivot Tables. Summarise your data by percentages.

Education vs Personal Loan

TD Account Vs Personal Loan

Online vs Personal Loan

Income_Category vs Personal Loan

EDUCATION VS PERSONAL LOAN

Count of Securities Accou	Personal Loan		
Education	No	Yes	Grand Total
Graduate	24%	4%	28%
Professional	26%	4%	30%
Undergraduate	40%	2%	42%
Grand Total	**90%**	**10%**	**100%**

ONLINE VS PERSONAL LOAN

Count of Personal Loan	Personal Loan		
Online	No	Yes	Grand Total
No	36.54%	3.78%	40.32%
Yes	53.86%	5.82%	59.68%
Grand Total	**90.40%**	**9.60%**	**100.00%**

TD ACCOUNT VS PERSONAL LOAN

Count of Personal Loan	Personal Loan		
TD Account	No	Yes	Grand Total
No	87.16%	6.80%	93.96%
Yes	3.24%	2.80%	6.04%
Grand Total	**90.40%**	**9.60%**	**100.00%**

INCOME_CATEGORY VS PERSONAL LOAN

Count of Personal Loan	Personal Loan		
Income Categorical	No	Yes	Grand Total
0-50	38.28%	0.00%	38.28%
100+	15.48%	8.76%	24.24%
51-100	36.64%	0.84%	37.48%
Grand Total	**90.40%**	**9.60%**	**100.00%**

The data suggests that customers with higher education level and TD account are more likely to avail personal loan.

Question 9:

Analyse the Pivot tables created in the previous question and state any anomaly that you observe. Which categorical variables appear most important for your further study if you want to analyse which customers are most likely to take personal loans and why?

Ans] Referring to the previous question it is visible that the variables named as Termed Deposit & Education are more important than the others. In addition to this, customers who have termed deposits are most likely to have a risk assurance factor against their termed deposits,so those customers have a higher chance of taking a personal loan. Second, the educational background of the customers will help us search for more targeted customers for loans. for eg, if the customer is a student, he would certainly not require a personal loan.

Question 10:

In the last campaign, bank reached out to 5000 customers out of which 480 customers accepted the personal loan offer. The bank incurred a huge cost in running a marketing campaign to reach out to so many customers. This is where you as a strategic business consultant step in. You are tasked to optimise the cost of this campaign by identifying the correct target base (without significant reduction in number of acceptance of offers). The bank can then send Personal Loan offers to these target customers who have a higher chance of accepting the offer. Based on your analysis, what strategy would you suggest to the management of HBFC bank?

Ans] With reference to the analysis and the dataset provided, the following were concluded:

1] Instead of directing all of our resources towards the entire population of customers, we must focus on a certain consumer group in order to effectively employ our resources and sell our goods to the public.

2] We should concentrate on graduates and professionals in considering the education criteria. Undergraduates make up the majority of customers when categorised according to education standards, yet their conversion rate is the lowest. Therefore, we ought to concentrate on professionals and recent grads.

3] Customers earning more than 100000k/year have maximum chances of buying our product.

4] During the analysis, it was observed that customers not using online banking tend to take more personal loans and therefore our focus should be on them